Available on the CERN CDS information server **CMS PAS EXO-09-012**

CMS Physics Analysis Summary

Contact: cms-pag-conveners-exotica@cern.ch 2009/07/30

Search for A Fourth Generation b' Quark in tW Final State at CMS in pp Collisions at $\sqrt{s} = 10$ TeV

The CMS Collaboration

Abstract

We present a feasibility study for the detection of a fourth generation b' quark pair production at $\sqrt{s} = 10$ TeV. If the mass of the b' quark is above the $m_t + m_W \approx 255$ GeV/c^2 threshold, the decay $b' \rightarrow tW$ is expected to be the dominant channel. By selecting same-sign dilepton and trilepton events, we estimate the sensitivity at LHC startup with the CMS detector. With a data set of 200 pb^{-1}, a $b'\overline{b}' \rightarrow t\bar{t}W^+W^-$ signal can be discovered with a significance of 3.7 standard deviations for $M(b') \sim$400 GeV/c^2. If no signal is observed in the data, a b' quark with a mass less than 485 GeV/c^2 can be excluded at the 95% confidence level.

1 Introduction

We present a study on the production heavy bottom-like fourth generation quark pairs, $b'\bar{b}'$ with the CMS detector. The mass of the b' quark is assumed to be above the top plus W threshold. The decay of $b'\bar{b}' \to tW^-\bar{t}W^+$ is expected to be dominant, hence a four W boson plus two b-jets final state is produced in the $pp \to b'\bar{b}'$ process. Each W boson can decay either leptonically ($W \to \ell\nu$) or hadronically ($W \to$ di-jet). Final states that contain same-sign dileptons or trileptons are rare in Standard Model processes, and hence the backgrounds are expected to be small. Hence this analysis is based on the same sign dilepton and trilepton (plus jets) signatures. A data set corresponding to 200 pb^{-1} integrated luminosity at 10 TeV is assumed for the search reach and exclusion limits. Three typical benchmark points are discussed in this analysis, with masses for the b's corresponding to 300, 400, and 500 GeV/c^2, which leads to production cross sections in pp collisions at $\sqrt{s} = 10$ TeV of 13.6 pb, 2.80 pb, and 0.78 pb, respectively. These cross sections are calculated with the Pythia [1] generator, at the leading-order (LO) in α_s. In this study, we use the Monte Carlo (MC) samples produced with a full or fast simulation of CMS detector.

2 Event Reconstruction and Selections

2.1 Trigger

The trigger system [2] at CMS consists of two steps, the Level-1 trigger and the High Level Trigger (HLT). The Level-1 trigger system is fast and uses trigger hardware based information, while the HLT system relies on the event reconstruction information. We require the events to pass either signal electron trigger or single muon trigger paths [1]. The electron path requires at least a electron candidate without any isolation requirement and a minimum E_T of 15 GeV, while the muon path requires a muon candidate associated with the muon system and tracker, and with a p_T threshold of 15 GeV/c.

2.2 Object Selections

Four types of objects are used in the analysis: electrons, muons, jets, and the transverse missing energy (MET). The electrons are built from a super cluster in the ECAL, and matched to the hits of the pixel detector. The momentum of a electron track is fitted using a Gaussian-Sum Filter algorithm (GSF) along its trajectory. The algorithm takes into account the possible emission of Bremsstrahlung photons in the silicon tracker layers. The transverse momentum of the electron candidates is required to be larger than 25 GeV/c, and the electron must be within the ECAL fiducial region ($|\eta| < 2.5$). The electron candidates in the ECAL barrel and ECAL endcap region is excluded from the analysis.

Electron candidates are cleaned further using the "robust-tight" electron identification selection criteria. These selection criteria imposes requirements on the ratio between the energy deposited in HCAL and in ECAL ($E_{\mathrm{HCAL}}/E_{\mathrm{ECAL}}$), the shower width in the η direction ($\sigma_{\eta\eta}$), the matching between the calorimeter shower and tracker in η and ϕ directions ($\Delta\eta(\mathrm{SC, track})$ and $\Delta\phi(\mathrm{SC, track})$). Table 1 summarizes these electron identification criteria for barrel and endcap regions. The track isolation variable is a sum of pT of charged tracks aligning within a cone of 0.3 radious. The centered electron is excluded by an inner cone of 0.015. The tracks have a minimum pT threshold of 1.0 GeV/c. We reject the electron candidates that have an isolation variable greater then 0.2 GeV/c.

[1] We select the "HLT_Ele15_SW_L1R" and "HLT_Mu15" trigger paths in the study.

Table 1: Summary of the "robust-tight" electron identification criteria.

Parameter	$E_{\mathrm{HCAL}}/E_{\mathrm{ECAL}}$	$\sigma_{\eta\eta}$	$\Delta\eta(\mathrm{SC, track})$	$\Delta\phi(\mathrm{SC, track})$
Criteria for barrel region	< 0.015	< 0.0092	< 0.0025	< 0.020
Criteria for endcap region	< 0.018	< 0.025	< 0.0040	< 0.020

Muons are reconstructed in three stages: the local reconstruction, stand-alone reconstruction and finally the global reconstruction. Starting from reconstructed hits, local track segments are reconstructed in the drift tubes and cathode strip chambers. Then the stand-alone muons are reconstructed using the full information from the muon system. The global muon reconstruction then further includes the hits at the silicon tracker. The transverse momentum of the muon candidates is required to be larger than 20 GeV/c, and to be in the central detector region ($|\eta| < 2.0$). In addition, the impact parameter of the muon relative to the beam spot position, $|d0|$, is required to be less then 2 mm. The number of hits in the tracker associated with the muon candidate must be greater or qual to 11. The muon identification "GlobalMuonPromptTight" is adopted in the analysis which is consistent with an upper limit on the normalized χ^2 of 10. We require a low activity of charged tracks by allowing maximally a p_T sum of 4 GeV/c from tracks within a cone of 0.3 radious, around the muon candidate, excluding the muon track itself. A minimum threshold on the electromagnetic activity is also required. A scalar sum of transverse energy depositions in ECAL in a cone radious of 0.3 should be less then 4 GeV as well.

Jets are reconstructed using the iterative cone algorithm with a cone radius of 0.5 in ΔR. To get the correct energy scale for the jets, the relative corrections which produce a uniform jet response along η and the absolute corrections which correct the jet energy back to the generator particle level are applied. We require a minimum p_T of 35 GeV/c, and the jet candidates to be in the region of $|\eta| < 2.3$.

The value of MET is determined from the calorimetric measurements and is corrected usign the jet energy scales. A correction for the missing energy contributed of the muon candidates is applied. We do not require a MET threshold in the analysis.

2.3 Event Selections and Expected Yields

For the same-sign dileptonic channel, exactly two leptons (either electron or muon) with the same electric charges ($\ell^+\ell^+$ or $\ell^-\ell^-$) are required. For the trileptonic channel, three leptons with the charge combinations of eg. $\ell^+\ell^+\ell^-$ or $\ell^+\ell^-\ell^-$ are required. The events are further required to have at least four or more jets in the final state for the same-sign dileptonic channel; for the trileptonic channel, at least two or more jets are required. In addition to these criteria on the number of lepton and jet candidates, at least a hard lepton ($p_T > 35$ GeV/c) and at least a hard jet ($p_T > 85$ GeV/c) are required in the events. We require a lepton-jet separation ($\Delta R(\mathrm{lepton, jet}) > 0.3$) in order to suppress the additional leptons from jets. Also, a lepton-lepton isolation ($\Delta R(\mathrm{lepton, lepton}) > 0.3$) is required to reject background from doubly reconstructed muon and electron (a photon is radiated from the muon, and is reconstructed as an electron candidate using the same charge track). An event is reject if the invariant mass of two muons or electrons of any charge is within the Z-boson mass region ($|M(\ell\ell) - M(Z)| < 10$ GeV/c^2). The above selection criteria are optimized assuming a $b' \to tW$ signal at 400 GeV/c^2.

The jet multiplicities for same-sign dileptonic and trileptonic channels are shown separately in Figure 1. The efficiencies for the signal are 1.25%, 1.89%, and 2.26% for a b' mass of 300, 400, and 500 GeV/c^2, including the branching ratios. The p_T distributions of the leading lepton and the leading jet, the MET distribution, and the distribution of the invariant mass of dileptons (with opposite charged) are shown in Figure 2. The vertical dashed lines in these plots represent the final selection criteria in the analysis. Signal events can be characterised with the observable, H_T, which is defined as following:

$$H_T = \sum p_T(\text{jets}) + \sum p_T(\text{leptons}) + \text{MET} . \tag{1}$$

The variable H_T is the most effective one that carries the information of b' mass, since the b' candidates cannot be completely reconstructed in this analysis. The expected distributions for H_T for 200 pb^{-1} of data are shown in Fig 3. The signal is very clear for the lower mass b' ($300 \sim 400\,\text{GeV}/c^2$). The expected signal and background yields with the final selection criteria are summarized in Table 2.

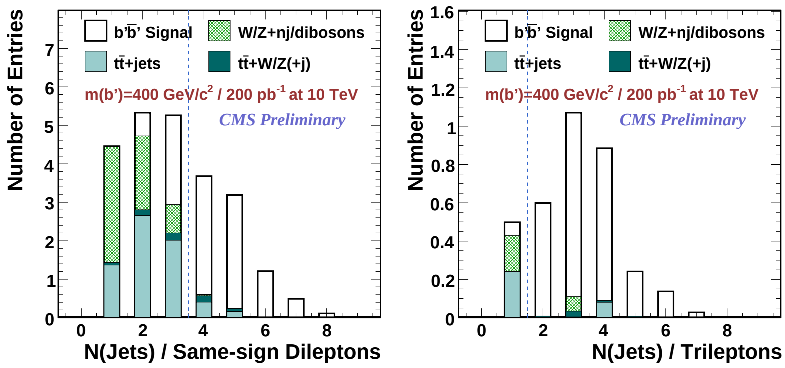

Figure 1: The jet multiplicities (jet counting) in the same-sign dileptonic channel (left) and in the trileptonic channel (right). The open area is the signal contributions for $400\,\text{GeV}/c^2$ b' pairs. The lighter (darker) shaded area shows the contributions from $t\bar{t}$+jets ($t\bar{t} + Z/W(+j)$) events. The hatched area represents the electroweak processes (W+jets, Z+jets, dibosons). The vertical dashed lines in the distributions are the selected minimum number of jets in the analysis.

3 Data Driven Signal Extraction

The background to the signal is expected to be small, but a data driven method is introduced for the background estimation. Control regions which are background rich are defined to normalize the background contributions. Events with two leptons of opposite charges are selected and all the other selection criteria are the same as for the signal regions. Figure 4 shows the distributions of H_T, $M(\ell^+\ell^-)$, and $N(\text{jets})$ for the events with opposite sign dileptons. If at least two jets are required in the events it will help to cancel the systematics on the jet counting for the trilepton channel (which also requires $N(\text{jets}) \geq 2$). For the consideration of same-sign dilepton channel, the required minimum number of jets is four, in order to match the same $N(\text{jets})$ requirement. The $t\bar{t}$ background is expected to dominate in these regions. The requirement of the Z-veto will reduce the contributions from the electroweak processes.

An iterative method is used to extract the signal yield. The number of background (N_B) and number of signal (N_S) events in the signal region can be extracted by

$$N_B = N_B^{\text{control}} \times R_B = (N^{\text{control}} - N_S^{\text{control}}) \times R_B \, , \tag{2}$$

$$N_S = N - N_B \, , \tag{3}$$

$$N_S^{\text{control}} = N_S / R_S \, , \tag{4}$$

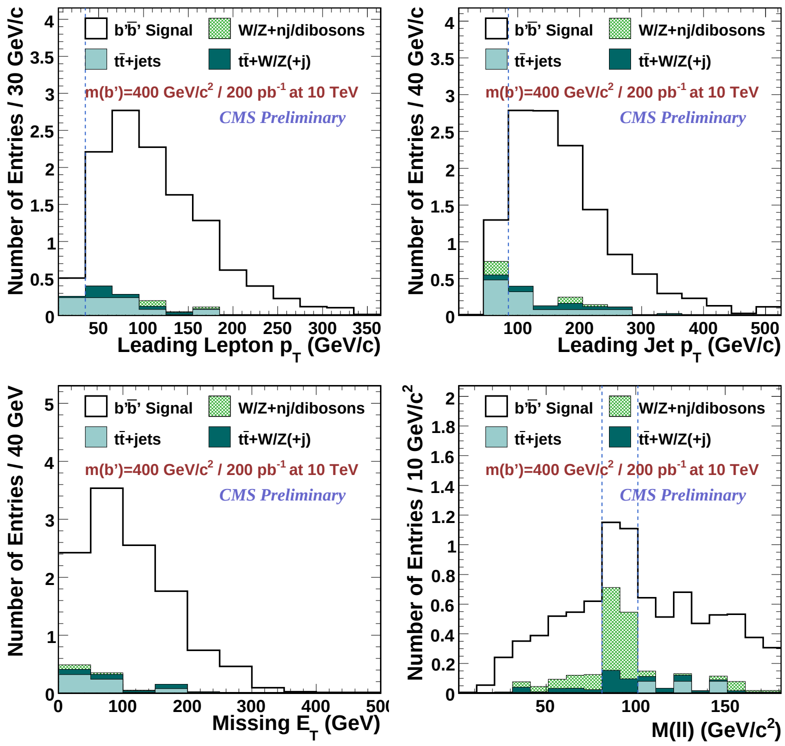

Figure 2: The distributions of transverse momentum p_T for the leading lepton, the leading jet, the transverse missing E_T (MET), and the invariant mass of two muons or electrons of any charge, $M(\ell\ell)$. The open area represents the signal contributions assuming a b' of 400 GeV$/c^2$. The lighter (darker) shaded area shows the contributions from $t\bar{t}$+jets ($t\bar{t} + Z/W(+j)$) events. The hatched area is occupied for the electroweak processes (W+jets, Z+jets, dibosons). The vertical dashed line on the MET is the required minimum value in the analysis; the vertical dashed lines on $M(\ell\ell)$ indicate the criteria for Z rejection.

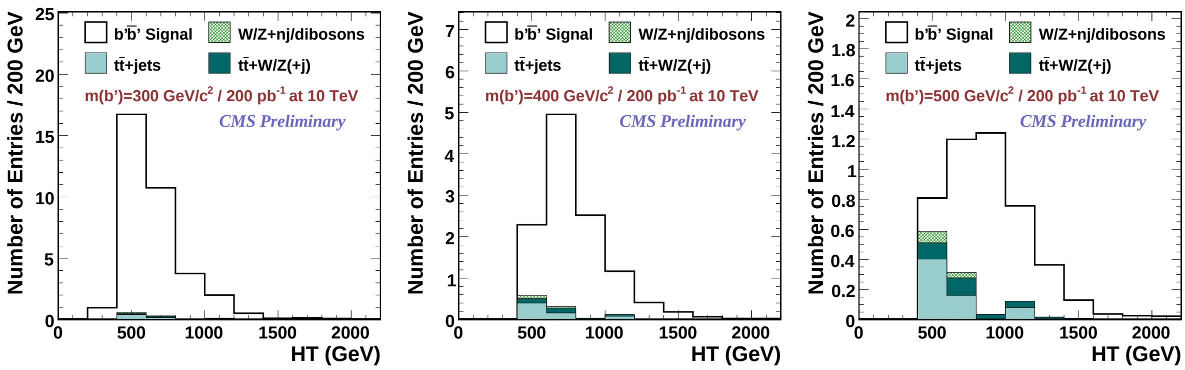

Figure 3: The distributions of H_T for three signal b' mass assumptions, both same-sign dileptonic channel and trileptonic channel are included. The open histograms are the expected signal area for 300 GeV$/c^2$ b' (left), 400 GeV$/c^2$ b' (middle), and 500 GeV$/c^2$ b' (right). The lighter shaded area represents $t\bar{t}$ plus jets background, while the darker shaded area shows the contributions from $t\bar{t} + Z/W(+j)$ processes. The hatched area represents the electroweak processes (W+jets, Z+jets, dibosons).

Table 2: Summary of expected signal and background yields normalized to 200 pb^{-1} integrated luminosity.

Process	Yield Same-sign 2L	3L	Sum
$b'\bar{b}'$, $M(b') = 300$ GeV$/c^2$	25.52	8.56	34.08
$b'\bar{b}'$, $M(b') = 400$ GeV$/c^2$	7.84	2.75	10.58
$b'\bar{b}'$, $M(b') = 500$ GeV$/c^2$	2.72	0.79	3.52
$t\bar{t}$+jets	0.56	0.08	0.64
$t\bar{t}W(+j)$	0.17	0.03	0.20
$t\bar{t}Z(+j)$	0.10	0.02	0.12
$t\bar{t}W^+W^-$	0.007	0.004	0.010
W+jets	< 0.08	< 0.08	< 0.08
Z+jets	< 0.08	0.08	0.08
WW	< 0.07	< 0.07	< 0.07
ZZ	< 0.01	< 0.01	< 0.01
WZ	0.03	< 0.03	0.03
Same-sign $WW + jj$	0.01	< 0.01	0.01
Background Sum	0.87	0.22	1.08

where N and N^{control} denote the total observed events in the signal region and in the control region. The ratio R_B (R_S) is the background (signal) ratio between the signal region and control region, and is obtained from MC simulations. The MC determined background ratio R_B is the major concern of the systematic uncertainty. Assuming no signal in the background region one will find still an excess in the signal region, be it lower than the true excess. Reiterating the control region assuming now a 'signal contamination' allows to recalculate the background estimation from the control region. By updating the signal and background yields in the control region (N_S^{control} and N_B^{control}) according to the equations above, a converging results on N_S is reached in few iterations. This number agrees with the value from the input MC.

4 Uncertainties

We perform an ensemble tests to extract the statistical uncertainties. The total yields in the signal region and in the control region are generated according to Poisson distributions (with the mean values set to the expected total yields). Assuming the generated N and N^{control} are the observed yields in the data, the signal and background yields are calculated using the method mentioned above. From the ensemble tests, the expected statistical uncertainties for 200 pb^{-1} data are estimated to be 17.7%, 33.3%, and 61.3% for a 300, 400, and 500 GeV$/c^2$ b' signal, respectively.

Systematic uncertainties are calculated for a cross-section measurement. The cross section is derived using the following equation

$$\sigma = \frac{N_S}{\epsilon_S \times \mathcal{L}}\,, \tag{5}$$

where ϵ_S represents the detector acceptance for the signal events, $\mathcal{L}$ is the integrated luminosity, and N_S is the signal yield calculated by the data driven method discussed in the previous section.

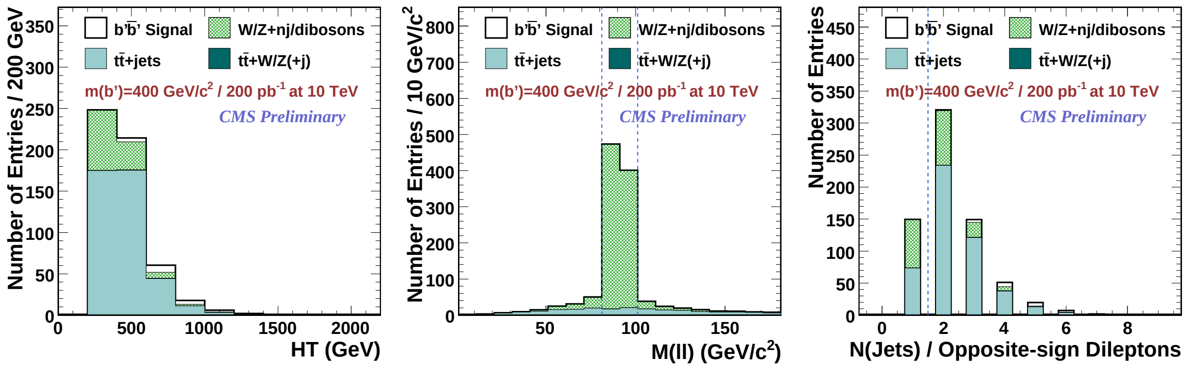

Figure 4: The distributions of H_T (left), dilepton mass $M(\ell^+\ell^-)$ (middle), and number of jets $N(\text{jets})$ (right) for the events with opposite sign dileptons. In the plots of H_T and $M(\ell\ell)$, at least two jets are required. For the plots of H_T and $N(\text{jets})$, the events in the Z mass region are rejected. The open area represents the signal contributions assuming a b' of 400 GeV$/c^2$. The lighter (darker) shaded area shows the contributions from $t\bar{t}$+jets ($t\bar{t} + Z/W(+j)$) events. The hatched area is occupied for the electroweak processes (W+jets, Z+jets, dibosons).

We first calculate the expected signal and background events based on a normal condition (N_S^{normal} and N_B^{normal}). We take the observed number of candidates in the signal region is exactly equal to $N^{\text{normal}} = N_S^{\text{normal}} + N_B^{\text{normal}}$. The systematic sources are tested one-by-one, and the parameters (N_S, ϵ_S, and $\mathcal{L}$) used in the cross-section calculation are changed accordingly. The variations on the cross-section are summed in quadrature for the final systematic uncertainty.

The dominated background source in this analysis is $t\bar{t}$+jets. These events usually come with a lepton with charge mis-identification, a fake lepton from jet, or a true lepton from b-jet. The major concern for the systematic uncertainties is on the background ratios (R_B) which are governed by the probability to observe a non-prompt lepton (lepton with charged mis-identification or lepton from b-jet) or a fake lepton in an event. Currently these ratios are only obtained from MC simulations, thus a conservative uncertainty on R_B (times 0.5 or 2) is taken. This uncertainty could be reduced with a futher data driven method. The possibility of charge mis-identification can be examined using Z-tagged events, which is discussed latter in this section. Other sources (fake leptons or lepton from b-jet) can be examined using other control samples. Another way is to make a comparison between the opposite dilepton and same-sign dilepton events with the same selection criteria as the signal and control region except for requiring less jets ($N(\text{jets}) \leq 1$). This determination could help to reduce the error on the background ratios, but relies more on the MC to go from the few jet control region to the multi-jet signal region.

The uncertainties on the cross section measurements are summarized in Table 4. The following contributions are considered in this study:

- Integrated luminosity: $\pm 10\%$ in the normalization.
- Non-prompt and fake leptons: vary the background ratios from half to double.
- Background cross sections: $t\bar{t}$+jets ($\pm 10\%$), $t\bar{t}$+bosons ($\pm 20\%$), di-bosons ($\pm 20\%$), $+100\%$ for the QCD multi-jets process.
- Jet energy scale: $\pm 10\%$ per jet.
- Jet reconstruction efficiency: assign 10% of the events to have an additional jet or to lose one jet.

- Pile-up jets: a comparison between the samples with and without pile-up simulations.
- Lepton identification, isolation, and trigger: $\pm 1\%$ per lepton
- Parton distribution function (PDF): re-weighting the events according to the uncertainty PDF sets [3].
- Monte Carlo statistics.

Table 3: Summary of systematic uncertainties on the cross-section measurements

$M(b')$ (GeV/c^2)	300	400	500
Integrated luminosity	11.1%	11.1%	11.1%
Non-prompt and fake leptons	$^{+2.9\%}_{-5.7\%}$	$^{+6.5\%}_{-13.0\%}$	$^{+17.1\%}_{-33.7\%}$
Background cross sections	0.3%	1.2%	3.6%
Jet energy scale	20.0%	8.4%	14.2%
Jet efficiency	4.2%	3.5%	5.2%
Leptons	2.3%	2.3%	2.3%
Pile-ups	0.5%	0.6%	1.0%
PDF	6.6%	8.9%	11.3%
MC Statistics	3.9%	4.4%	8.3%
Statistics (200 pb^{-1})	17.7%	33.3%	61.3%
Sum: systematics	$^{+25\%}_{-25\%}$	$^{+19\%}_{-22\%}$	$^{+29\%}_{-41\%}$
Sum: syst. + stat.	$^{+30\%}_{-31\%}$	$^{+38\%}_{-40\%}$	$^{+68\%}_{-74\%}$

5 Significance Estimation

Several significance estimators are discussed in Ref. [4]:

$$S_{12} = 2\left(\sqrt{N_S + N_B} - \sqrt{N_B}\right) \times \sqrt{\frac{N_B}{N_B + \Delta_B^2}}\,, \tag{6}$$

$$S_{cP} : \int_{-\infty}^{\infty} \sum_{N=N_S+N_B}^{\infty} P(N;\mu)\, G(\mu; N_B, \Delta_B)\, d\mu = \int_{S_{cP}}^{\infty} G(x;1,0)\, dx\,, \tag{7}$$

where the $P(N;\mu)$ is the Poisson probability distribution function with mean value μ, and $G(x;\bar{x},\sigma)$ is the Gaussian function with a mean value of $\bar{x}$ and the standard deviation σ. The S_{cP} variable is recommended for a counting experiment.

As summarized in Table 5, total systematic uncertainty on background yield is estimated to be $^{+114\%}_{-67\%}$ (that is, $N_B = 1.08^{+1.23}_{-0.72}$ events), mainly contributed by the uncertainty on the background ratios. This uncertainty is included in the significance calculations. The results are shown in Table 6. With an initial data set of 200 pb^{-1} at 10 TeV, we are already able to observe a $b'\overline{b}'$ signal with a significance of 3.7σ if the b' mass is around 400 GeV/c^2. The exclusion limits on the $pp \to b'\overline{b}'$ cross sections are estimated for a null hypothesis using Bayesian statistics. As shown in Figure 5, using LO cross sections predictions as provided by Pythia, we are able to exclude the productions of $pp \to b'\overline{b}'$ up to a b' mass of 485 (405) GeV/c^2 at the 95% confidence level with a data set of 200 (60) pb^{-1}.

Table 4: Summary of systematic uncertainties on the cross-section measurements

$M(b')$ (GeV/c^2)	300	400	500
Integrated luminosity	11.1%	11.1%	11.1%
Non-prompt and fake leptons	$^{+2.9\%}_{-5.7\%}$	$^{+6.5\%}_{-13.0\%}$	$^{+17.1\%}_{-33.7\%}$
Background cross sections	0.3%	1.2%	3.6%
Jet energy scale	20.0%	8.4%	14.2%
Jet efficiency	4.2%	3.5%	5.2%
Leptons	2.3%	2.3%	2.3%
Pile-ups	0.5%	0.6%	1.0%
PDF	6.6%	8.9%	11.3%
MC Statistics	3.9%	4.4%	8.3%
Statistics (200 pb^{-1})	17.7%	33.3%	61.3%
Sum: systematics	$^{+25\%}_{-25\%}$	$^{+19\%}_{-22\%}$	$^{+29\%}_{-41\%}$
Sum: syst. + stat.	$^{+30\%}_{-31\%}$	$^{+38\%}_{-40\%}$	$^{+68\%}_{-74\%}$

Table 5: Summary of systematic uncertainties on the background yields (ΔB). The error from integrated luminosity is totally canceled.

	$\Delta B/B$
Integrated luminosity	0.0%
Non-prompt and fake leptons	$^{+105.0\%}_{-50.6\%}$
Background cross sections	11.7%
Jet energy scale	32.0%
Jet efficiency	11.0%
Leptons	0.2%
Pile-ups	2.3%
PDF	1.2%
MC Statistics	24.8%
Sum	$^{+114\%}_{-67\%}$

6 Conclusion

In conclusion, we present a first study on the production of bottom-like fourth generation quark pairs at LHC. Based on the Monte Carlo simulations with the CMS detector and for a data sample of 200 pb^{-1} integrated luminosity at 10 TeV, we are able to observe the process $pp \to b'\bar{b}' \to t\bar{t}W^+W^-$ with a significance of 3.7 standard deviations, if the mass of the b' quark is around 400 GeV/c^2 and the production cross section is 2.80 pb. The total uncertainty on the cross section measurement for such b' signal is around 40%. If the observed events are consistent with Standard Model background processes only, a b' quark with mass less than 485 GeV/c^2 can be excluded at the 95% confidence level.

References

[1] T. Sjostrand et al., "High-energy-physics event generation with PYTHIA 6.1," *Comput. Phys. Commun.* **135** (2001) 238–259, arXiv:hep-ph/0010017. doi:10.1016/S0010-4655(00)00236-8.

Table 6: Expected significance for 200 pb^{-1} data for an assumption of 300 GeV/c^2, 400 GeV/c^2, and 500 GeV/c^2 b' signal. Since the background uncertainty is larger then the expected signal yield for 500 GeV/c^2 b', the values of S_{12} and S_{cP} are effectively zero accord to the estimator.

b' Mass	300 GeV/c^2	400 GeV/c^2	500 GeV/c^2
$b'\bar{b}'$ LO cross section	13.6 pb	2.80 pb	0.78 pb
Expected signal yield	34.08	10.58	3.52
Expected background yield	$1.08^{+1.23}_{-0.72}$		
S_{12}	6.3σ	3.1σ	1.4σ
S_{cP}	9.0σ	3.7σ	1.4σ

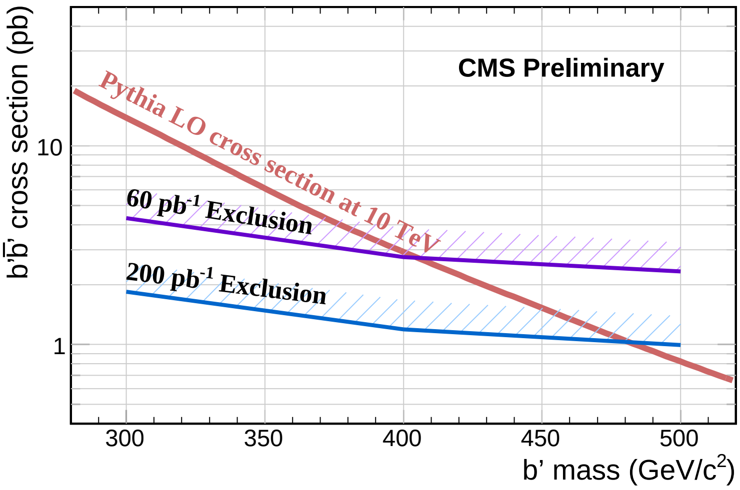

Figure 5: The exclusion limits at the 95% confidence level on the $pp \rightarrow b'\bar{b}'$ production cross sections. Comparing to the Pythia LO cross sections, b' masses less than 485 (405) GeV/c^2 can be excluded with a data set of 200 (60) pb^{-1} for a null hypothesis.

[2] **CMS** Collaboration, "CMS High Level Trigger," *CERN/LHCC* **2007/021** (2007).

[3] M. R. Whalley, D. Bourilkov, and R. C. Group, "The Les Houches Accord PDFs (LHAPDF) and Lhaglue," `arXiv:hep-ph/0508110`.

[4] S. I. Bityukov and N. V. Krasnikov, "Uncertainties and discovery potential in planned experiments," `arXiv:hep-ph/0204326`.